# Part II

# Discrete Random Variables

In Part I, we saw that experiments are classified as either having a discrete sample space, with a countable number of possible outcomes, or a continuous sample space, with an uncountable number of possible outcomes. In this part, our focus will be on the discrete world. In Part III we will focus on the continuous world.

We start, in Chapter 3, by introducing the notion of a discrete random variable. We then show that everything that we've learned about probability on events applies to random variables as well. In this chapter, we cover the most common discrete distributions: the Bernoulli, Binomial, Geometric, and Poisson.

Chapter 4 is devoted to understanding expectation of discrete random variables. This includes linearity of expectation and conditional expectation. We end with a discussion of Simpson's paradox.

In Chapter 5, we move on to variance and higher moments of discrete random variables. We also introduce the notion of a sum of random variables, where the number being summed is itself a random variable. We next turn to the tail of a random variable, namely the probability that the random variable exceeds some value, introducing some very simple tail bounds, as well as the concept of stochastic dominance. We end with a discussion of the inspection paradox, which is one of the more subtle consequences of high variability.

Finally, in Chapter 6, we finish off our unit on discrete random variables by introducing the z-transform, a moment-generating function which is tailored for discrete random variables. The z-transform allows us to quickly compute all higher moments of random variables. It also has many other applications, including solving recurrence relations.

# 3 Common Discrete Random Variables

While the previous chapter covered probability on events, in this chapter we will switch to talking about random variables and their corresponding distributions. We will cover the most common discrete distributions, define the notion of a joint distribution, and finish with some practical examples of how to reason about the probability that one device will fail before another.

## 3.1 Random Variables

Consider an experiment, such as rolling two dice. Suppose that we are interested in the sum of the two rolls. That sum could range anywhere from 2 to 12, with each of these events having a different probability. A *random variable*, $X$, associated with this experiment is a way to represent the value of the experiment (in this case the sum of the rolls). Specifically, when we write $X$, it is understood that $X$ has many instances, ranging from 2 to 12 and that different instances occur with different probabilities. For example, $\mathbf{P}\{X = 3\} = \frac{2}{36}$.

Formally, we say,

**Definition 3.1** *A* **random variable** *(r.v.) is a real-valued function of the outcome of an experiment involving randomness.*

For the above experiment, r.v. $X$ could be the sum of the rolls, while r.v. $Y$ could be the sum of the squares of the two rolls, and r.v. $Z$ could be the value of the first roll only. Any real-valued function of the outcome is legitimate.

As another experiment, we can imagine throwing two darts at the interval $[0, 1]$, where each dart is equally likely to land anywhere in the interval. Random variable $D$ could then represent the distance between the two darts, while r.v. $L$ represents the position of the leftmost dart.

**Definition 3.2** *A* **discrete random variable** *can take on at most a countably infinite number of possible values, whereas a* **continuous random variable** *can take on an uncountable set of possible values.*

**Question:** Which of these random variables is discrete and which is continuous?

(a) The sum of the rolls of two dice
(b) The number of arrivals at a website by time $t$
(c) The time until the next arrival at a website
(d) The CPU requirement of an HTTP request

**Answer:** The sum of rolls can take on only a finite number of values – those between 2 and 12 – so it clearly is a discrete r.v. The number of arrivals at a website can take on the values: $0, 1, 2, 3, \ldots$ namely a countable set; hence this is discrete as well. Time, in general, is modeled as a continuous quantity, even though there is a non-zero granularity in our ability to measure time via a computer. Thus quantities (c) and (d) are continuous random variables.

*We use capital letters to denote random variables*. For example, $X$ could be a r.v. denoting the sum of two dice, where

$$\mathbf{P}\{X = 7\} = \mathbf{P}\{(1,6) \text{ or } (2,5) \text{ or } (3,4), \ldots, \text{ or } (6,1)\} = \frac{1}{6}.$$

**Key insight: Because the "outcome of the experiment" is just an event, all the theorems that we learned about events apply to random variables as well.** For example, $X = 7$ above is an event. In particular, the Law of Total Probability (Theorem 2.18) holds. For example, if $N$ denotes the number of arrivals at a website by time $t$, then $N > 10$ is an event. We can then use conditioning on events to get

$$\mathbf{P}\{N > 10\} = \mathbf{P}\{N > 10 \mid \text{ weekday }\} \cdot \frac{5}{7} + \mathbf{P}\{N > 10 \mid \text{ weekend }\} \cdot \frac{2}{7}.$$

All of this will become more concrete when we study examples of random variables.

## 3.2 Common Discrete Random Variables

Discrete random variables take on a countable number of values, each with some probability. A discrete r.v. is associated with a discrete probability distribution that represents the likelihood of each of these values occurring. We will sometimes go so far as to define a r.v. by the distribution associated with it, omitting the whole discussion of an "experiment."

**Definition 3.3** *Let X be a discrete r.v. Then the* **probability mass function (p.m.f.)**, $p_X(\cdot)$ *of X, is defined as:*

$$p_X(a) = \mathbf{P}\{X = a\} \qquad \textit{where} \qquad \sum_x p_X(x) = 1.$$

*The* **cumulative distribution function** *of X is defined as:*

$$F_X(a) = \mathbf{P}\{X \le a\} = \sum_{x \le a} p_X(x).$$

*The* **tail** *of X is defined as:*

$$\overline{F}_X(a) = \mathbf{P}\{X > a\} = \sum_{x > a} p_X(x) = 1 - F_X(a).$$

Common discrete distributions include the Bernoulli, the Binomial, the Geometric, and the Poisson, all of which are discussed next.

### 3.2.1 The Bernoulli($p$) Random Variable

Consider an experiment involving a single coin flip, where the coin has probability $p$ of coming up heads and $1 - p$ of coming up tails.

Let r.v. $X$ represent the outcome of the experiment, that is, the value of the coin. We say that the value is 1 if the coin comes up heads and 0 otherwise. Then,

$$X = \begin{cases} 1 & \text{w/ prob } p \\ 0 & \text{otherwise} \end{cases}.$$

We say that $X$ is a r.v. *drawn from* the Bernoulli($p$) distribution, and we write:

$$X \sim \text{Bernoulli}(p).$$

The p.m.f. of r.v. $X$ is defined as follows:

$$p_X(1) = p$$
$$p_X(0) = 1 - p.$$

The p.m.f. is depicted in Figure 3.1.

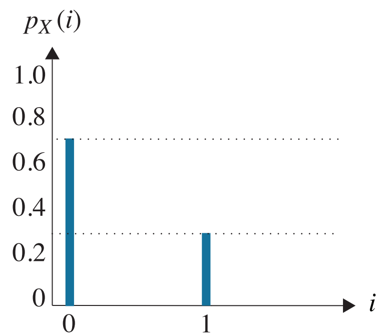


**Figure 3.1** *Probability mass function of the Bernoulli$(p = 0.3)$ distribution.*

### 3.2.2 The Binomial($n$,$p$) Random Variable

Now consider an experiment where we again have a coin with probability $p$ of coming up heads (success). This time we flip the coin $n$ times (these are independent flips).

Let r.v. $X$ represent the number of heads (successes). Observe that $X$ can take on any of these (discrete) values: $0, 1, 2, \ldots, n$.

The p.m.f. of r.v. $X$ is defined as follows:

$$\begin{aligned} p_X(i) &= \mathbf{P}\{X = i\} \\ &= \binom{n}{i} p^i (1-p)^{n-i}, \text{ where } i = 0, 1, 2, \ldots, n. \end{aligned}$$

A r.v. $X$ with the above p.m.f. is said to be drawn from the Binomial$(n, p)$ distribution, written: $X \sim$ Binomial$(n, p)$. The p.m.f. is shown in Figure 3.2.

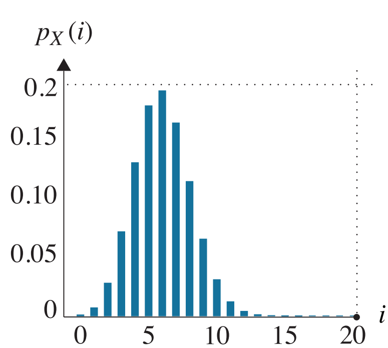


**Figure 3.2** *Probability mass function of the Binomial$(n = 20, p = 0.3)$ distribution.*

Observe that the sum of the p.m.f. is 1, as desired:

$$\sum_{i=0}^{n} p_X(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1. \quad \checkmark$$

Here we've used the binomial expansion from Section 1.5.

### 3.2.3 The Geometric($p$) Random Variable

Again consider an experiment where we have a coin with probability $p$ of coming up heads (success). We now flip the coin until we get a success; these are independent trials, each distributed Bernoulli($p$).

Let r.v. $X$ represent the number of flips until we get a success.

The p.m.f. of $X$ is defined as follows:

$$\begin{aligned} p_X(i) &= \mathbf{P}\{X = i\} \\ &= (1-p)^{i-1} p, \text{ where } i = 1, 2, 3, \ldots. \end{aligned}$$

A r.v. $X$ with the above p.m.f. is said to be drawn from the Geometric($p$) distribution, written: $X \sim$ Geometric($p$). The p.m.f. is shown in Figure 3.3.

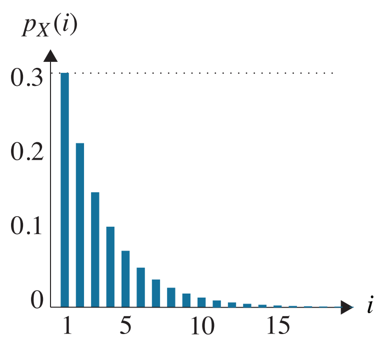


**Figure 3.3** *Probability mass function of the Geometric($p = 0.3$) distribution.*

**Question:** What is $\overline{F}_X(i)$?

**Answer:**

$$\overline{F}_X(i) = \mathbf{P}\{X > i\} = \mathbf{P}\{\text{First } i \text{ flips were tails}\} = (1-p)^i.$$

Observe that the sum of the p.m.f. is 1, as desired:

$$\sum_{i=1}^{\infty} p_X(i) = \sum_{i=1}^{\infty} (1-p)^{i-1} \cdot p = \sum_{i=0}^{\infty} (1-p)^i \cdot p = p \cdot \frac{1}{1-(1-p)} = 1. \quad \checkmark$$

Here we've used the Geometric series sum from Section 1.1.

**Question:** Let's review. Suppose you have a room of $n$ disks. Each disk independently dies with probability $p$ each year. How are the following quantities distributed?

(a) The number of disks that die in the first year
(b) The number of years until a particular disk dies
(c) The state of a particular disk after one year

**Answer:** The distributions are: (a) Binomial$(n, p)$; (b) Geometric$(p)$; (c) Bernoulli$(p)$.

### 3.2.4 The Poisson($\lambda$) Random Variable

We define the Poisson($\lambda$) distribution via its p.m.f. Although the p.m.f. does not appear to have any meaning at present, we will see that it comes up in many applications where the distribution is bell-shaped, but has a lower bound of 0. For example, we will see in Chapter 5 that the Poisson distribution is a good representation of the number of pairs of shoes owned by people. Likewise, in Chapter 12, we will see that the Poisson distribution occurs naturally when looking at a mixture of a very large number of independent sources, each with a very small individual probability. It can therefore be a reasonable approximation for the distribution of the number of arrivals to a website or a router.

If $X \sim$ Poisson$(\lambda)$, then

$$p_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}, \text{ where } i = 0, 1, 2, \ldots.$$

The p.m.f. for the Poisson($\lambda$) distribution is shown in Figure 3.4.

The sum of the p.m.f. is again 1, as desired:

$$\sum_{i=0}^{\infty} p_X(i) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} \cdot e^{\lambda} = 1. \quad \checkmark$$

Here we've used the Taylor series expansion from (1.11) of Section 1.4.

**Question:** Does the shape of the Poisson distribution remind you of other distributions?

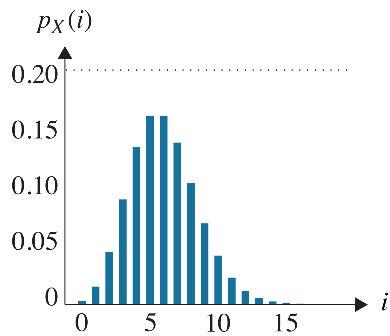


**Figure 3.4** *Probability mass function of the Poisson($\lambda = 6$) distribution.*

**Answer:** The Poisson distribution does not look all that different from the Binomial distribution. It too has a bell-like shape. However, it has an infinite range. In Exercise 3.8 we will see that if $n$ is large and $p$ is small, then Binomial$(n, p)$ is actually very close to Poisson$(np)$. The Poisson distribution is also similar to the Normal distribution (Chapter 9), except that it is lower-bounded by 0.

## 3.3 Multiple Random Variables and Joint Probabilities

We are often interested in probability statements concerning two or more random variables simultaneously. For example, imagine that we have $n$ disks, each of which fails with probability $p$ every day. We might want to know the probability that all $n$ disks fail on the same day, or the probability that disk 1 fails before disk 2. In asking such questions, we often are assuming that the failure of disks is **independent** (in which case we often say that the disks "independently fail" with probability $p$ on each day). By independent, we mean that the fact that one disk fails doesn't influence the failure of the other disks. However, it could be that the failures are **positively correlated**. By this we mean that the fact that one disk fails makes it more likely that other disks fail as well (for example, maybe the fact that a disk failed means there are mice in the building, which in turn can influence other disks).

In the above scenario, the state of each disk (working or failed) is a r.v. There are several ways to reason about multiple random variables. We introduce two techniques in this section. The first technique involves using the joint p.m.f. and is illustrated in Example 3.6. The second involves conditioning one r.v. on another and is illustrated in Example 3.8.

**Definition 3.4** *The* **joint probability mass function** *between discrete random variables X and Y is defined by*

$$p_{X,Y}(x, y) = \mathbf{P}\{X = x \;\&\; Y = y\}.$$

*This is equivalently written as* $\mathbf{P}\{X = x, Y = y\}$ *or as* $\mathbf{P}\{X = x \cap Y = y\}$. *By definition:*

$$\sum_x \sum_y p_{X,Y}(x, y) = 1.$$

**Question:** What is the relationship between $p_X(x)$ and $p_{X,Y}(x, y)$?

**Answer:** Via the Law of Total Probability, we have:

$$p_X(x) = \sum_y p_{X,Y}(x, y) \quad \text{and} \quad p_Y(y) = \sum_x p_{X,Y}(x, y).$$

When written this way, $p_X(x)$ is often referred to as the **marginal probability mass function** of $X$. The term "marginal" comes from the fact that $p_X(x)$ here would appear in the margins of a joint p.m.f. table, after summing an entire column over all $y$ values.

Similarly to the way we defined two events $E$ and $F$ as being independent, we can likewise define two random variables as being independent. This is because $X = x$ and $Y = y$ are events.

**Definition 3.5** *We say that discrete random variables X and Y are* **independent**, *written* $X \perp Y$, *if*

$$\mathbf{P}\{X = x \;\&\; Y = y\} = \mathbf{P}\{X = x\} \cdot \mathbf{P}\{Y = y\}, \quad \forall x, y$$

*or, equivalently,*

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y).$$

**Question:** If $X$ and $Y$ are independent, what does this say about $\mathbf{P}\{X = x \mid Y = y\}$?

**Answer:** Again, since $X = x$ and $Y = y$ are events, we can apply the simple conditioning formula that we learned in Chapter 2. As expected,

$$\mathbf{P}\{X = x \mid Y = y\} = \frac{\mathbf{P}\{X = x \;\&\; Y = y\}}{\mathbf{P}\{Y = y\}} = \frac{\mathbf{P}\{X = x\} \cdot \mathbf{P}\{Y = y\}}{\mathbf{P}\{Y = y\}} = \mathbf{P}\{X = x\}.$$

**Example 3.6 (Who fails first?)**

Here's a question that commonly comes up in industry, but isn't immediately obvious. You have a disk with probability $p_1$ of failing each day. You have a CPU which independently has probability $p_2$ of failing each day.

**Question:** What is the probability that your disk fails *before* your CPU?

Before you look at the answer, try to think for yourself what the answer might be. Is it $|p_1 - p_2|$, or $\frac{p_1}{p_2}$, or $p_1(1-p_2)$?

**Answer:** We model the problem by considering two Geometric random variables and deriving the probability that one is smaller than the other. Let $X_1 \sim \text{Geometric}(p_1)$ and $X_2 \sim \text{Geometric}(p_2)$, where $X_1 \perp X_2$. We want $\mathbf{P}\{X_1 < X_2\}$.

$$\begin{aligned}
\mathbf{P}\{X_1 < X_2\} &= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} p_{X_1,X_2}(k,k_2) \\
&= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} p_{X_1}(k) \cdot p_{X_2}(k_2) \qquad \text{(by independence)} \\
&= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} (1-p_1)^{k-1} p_1 \cdot (1-p_2)^{k_2-1} p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1} p_1 \sum_{k_2=k+1}^{\infty} (1-p_2)^{k_2-1} p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1} p_1 (1-p_2)^k \sum_{k_2=1}^{\infty} (1-p_2)^{k_2-1} p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1} p_1 (1-p_2)^k \cdot 1 \\
&= p_1(1-p_2) \sum_{k=1}^{\infty} [(1-p_2)(1-p_1)]^{k-1} \\
&= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}. \qquad (3.1)
\end{aligned}$$

**Question:** Explain why your final expression (3.1) makes sense.

**Answer:** Think about $X_1$ and $X_2$ in terms of coin flips. Notice that all the flips are irrelevant until the final flip, since before the final flip both the $X_1$ coin and the $X_2$ coin only yield tails. $\mathbf{P}\{X_1 < X_2\}$ is the probability that on that final flip, where by definition at least one coin comes up heads, it is the case that the $X_1$

coin is heads and the $X_2$ coin is tails. So we're looking for the probability that the $X_1$ coin produces a heads and the $X_2$ coin produces a tails, conditioned on the fact that they're not both tails, which is derived as:

$$\begin{aligned}\mathbf{P}\{\text{Coin } 1 = H \text{ \& Coin } 2 = T \mid \text{not both } T\} &= \frac{\mathbf{P}\{\text{Coin } 1 = H \text{ \& Coin } 2 = T\}}{\mathbf{P}\{\text{not both } T\}} \\ &= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}. \quad \checkmark\end{aligned}$$

Another way to approach Example 3.6 is to use conditioning. In computing the probability of an event, we saw in Chapter 2 that it is useful to condition on other events. We can use this same idea in computing probabilities involving random variables, because $X = k$ and $Y = y$ are just events. Thus, Theorem 3.7 follows immediately from the Law of Total Probability (Theorem 2.18).

**Theorem 3.7 (Law of Total Probability for Discrete R.V.)** *We can express the probability of an event $E$ by conditioning on a discrete r.v. $Y$ as follows:*

$$\mathbf{P}\{E\} = \sum_y \mathbf{P}\{E \cap Y = y\} = \sum_y \mathbf{P}\{E \mid Y = y\} \cdot \mathbf{P}\{Y = y\}.$$

*Likewise, for discrete random variables $X$ and $Y$, we can express the probability of the event $X = k$ by conditioning on the value of $Y$ as follows:*

$$\mathbf{P}\{X = k\} = \sum_y \mathbf{P}\{X = k \cap Y = y\} = \sum_y \mathbf{P}\{X = k \mid Y = y\} \cdot \mathbf{P}\{Y = y\}.$$

As always, being able to condition is a *huge* tool! It allows us to break a problem into a number of simpler problems. The trick, as usual, is knowing what to condition on.

**Example 3.8 (Who fails first, revisited)**

Suppose again that your disk has probability $p_1$ of failing each day, and your CPU independently has probability $p_2$ of failing each day.

**Question:** What is the probability that your disk fails *before* your CPU? This time use conditioning to determine this probability.

**Answer:** Again, let $X_1 \sim \text{Geometric}(p_1)$ and $X_2 \sim \text{Geometric}(p_2)$, where $X_1 \perp X_2$.

$$\begin{aligned}
\mathbf{P}\{X_1 < X_2\} &= \sum_{k=1}^{\infty} \mathbf{P}\{X_1 < X_2 \mid X_1 = k\} \cdot \mathbf{P}\{X_1 = k\} \\
&= \sum_{k=1}^{\infty} \mathbf{P}\{k < X_2 \mid X_1 = k\} \cdot \mathbf{P}\{X_1 = k\} \\
&= \sum_{k=1}^{\infty} \mathbf{P}\{X_2 > k\} \cdot \mathbf{P}\{X_1 = k\} \qquad \text{(by independence)} \\
&= \sum_{k=1}^{\infty} (1 - p_2)^k \cdot (1 - p_1)^{k-1} \cdot p_1 \\
&= p_1(1 - p_2) \sum_{k=1}^{\infty} [(1 - p_2)(1 - p_1)]^{k-1} \\
&= \frac{p_1(1 - p_2)}{1 - (1 - p_2)(1 - p_1)}.
\end{aligned}$$

Unsurprisingly, conditioning leads to a simpler solution.

## 3.4 Exercises

3.1 **ORs and ANDs**
Two fair coins are flipped. Let $X$ represent the logical OR of the two flips. Let $Y$ represent the logical AND of the two flips.
(a) What is the distribution of $X$?
(b) What is the distribution of $Y$?
(c) What is the distribution of $X + Y$?

3.2 **If at first you don't succeed**
Every day, independently at random, I win a prize with probability $\frac{1}{100}$. What is the probability that it takes more than 100 days to win a prize?

3.3 **Independence**
We're given a joint p.m.f. for two discrete random variables $X$ and $Y$.

| | $Y = 1$ | $Y = 2$ | $Y = 3$ |
|---|---|---|---|
| $X = 0$ | $1/8$ | $1/4$ | $1/8$ |
| $X = 1$ | $1/8$ | $0$ | $3/8$ |

(a) What is $p_{X,Y}(0, 1)$? What is $p_X(0)$? What is $p_Y(1)$?
(b) Are $X$ and $Y$ independent?

3.4 **From 10 disks to 1**
Today you have 10 working disks. Suppose that each disk independently dies with probability $p$ each day. What is the probability that tomorrow you have just 1 working disk?

3.5 **Independence of random variables**
Sachit has been studying the definition of independence of discrete random variables (Definition 3.5). He's wondering if the following statement is a corollary of the definition:

$$\text{If } X \perp Y, \text{ then } \mathbf{P}\{X > i \;\&\; Y > j\} = \mathbf{P}\{X > i\} \cdot \mathbf{P}\{Y > j\}.$$

Prove or disprove this statement.

3.6 **More independence practice**
We're given a joint p.m.f. for two random variables $X$ and $Y$.

| | $Y = 1$ | $Y = 2$ | $Y = 3$ |
|---|---|---|---|
| $X = 1$ | 3/8 | 3/16 | 1/4 |
| $X = 2$ | 1/8 | 1/16 | 0 |

(a) Are $X$ and $Y$ independent?
(b) What is $\mathbf{P}\{X = 1 \mid Y > 1\}$?
(c) Find an event $A$ where $X$ and $Y$ are conditionally independent given $A$.

3.7 **Sum of two independent Binomials**
Let $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(n, p)$, where $X \perp Y$. What is the distribution of $Z = X + Y$? [Hint: Don't try to do this via math. Think about the experiment.]

3.8 **Poisson approximation to Binomial**
You will prove that the Binomial$(n, p)$ distribution is well approximated by the Poisson$(np)$ distribution when $n$ is large and $p$ is small. Let $X \sim$ Binomial$(n, p)$ and consider $p_X(i)$, for an arbitrary fixed value of $i \geq 0$. In your expression for $p_X(i)$, set $p = \lambda/n$ so that $p_X(i)$ is expressed in terms of only $\lambda$ and $n$. Expand out all the "choose" terms. Now take the limit as $n \to \infty$, while remembering that $i$ is a fixed constant. Show that $p_X(i)$ approaches the p.m.f. of a Poisson$(\lambda)$ r.v.

3.9 **COVID testing**
[Proposed by Vanshika Chowdhary] On day 0, you take a long-distance flight and you are infected with COVID with probability $\frac{1}{2}$. Being a responsible citizen, you decide to quarantine for 14 days. You also visit a wizard, who gives you some special beans to help cure you, just in case you are sick. You take the beans every day, starting on day 1 of your quarantine. Each day, the beans have a $\frac{1}{8}$ chance of immediately curing you. Suppose

you are tested at the end of day 14 (after 14 days of taking beans) and the test comes back negative. What is the probability that you were actually infected with COVID on day 0? Assume that the test is fully accurate.

3.10 **Marginal probability**
An urn contains $n$ balls, which are numbered $1, 2, \ldots, n$. Suppose that we draw $k < n$ balls *without replacement* from the urn. Each ball is selected at random. Specifically, in the first draw, each ball has probability $\frac{1}{n}$ of being selected. In the second draw, each of the remaining $n-1$ balls has probability $\frac{1}{n-1}$ of being selected, and so on. Let $X_i$ denote the number on the $i$th ball drawn. Your goal is to prove that

$$\mathbf{P}\{X_i = \ell\} = \frac{1}{n}.$$

To do that, follow these steps:
(a) Are the $X_i$'s independent?
(b) Write an expression for $\mathbf{P}\{X_1 = a_1, X_2 = a_2, \ldots, X_k = a_k\}$, where $1 \le a_i \le n$.
(c) Express the marginal probability, $\mathbf{P}\{X_i = \ell\}$, as a sum

$$\mathbf{P}\{X_i = \ell\} = \sum \mathbf{P}\{X_1 = a_1, \ldots, X_{i-1} = a_{i-1}, X_i = \ell, \ldots, X_k = a_k\}.$$

What is the sum over?
(d) Evaluate the summation from (c). Start by evaluating the term inside the sum. Then determine the number of terms being summed.

3.11 **Binary symmetric channel (BSC)**
A binary symmetric channel is a communications model used in coding theory. There is a transmitter who wishes to send a bit, $B$. There is noise, $N$, which may corrupt the bit, and there is a final output $Y$, where

$$Y = B \oplus N.$$

Here, $\oplus$ is a binary sum. Assume that $B \sim \text{Bernoulli}(p)$ and that $N \sim \text{Bernoulli}(0.5)$ and that $N \perp B$. Can we say that $B \perp Y$? Why or why not?

3.12 **Noisy reading from flash storage**
Flash memories are a type of storage media which provide orders of magnitude faster access to data as compared to hard disks. However, one of the downsides of flash memories is that they are prone to error when reading. You have two flash memory devices, F1 and F2. The noisy readings from F1 and F2 are modeled as follows:
- F1: For any stored bit, the value read is flipped with probability $p_1$.
- F2: For any stored bit, the value read is flipped with probability $p_2$.

Suppose you write a bit into both F1 and F2 (i.e., the same bit is written into both devices), and that F1 and F2 act independently on that bit. A day later, you read the bit that you wrote from F1 and from F2. Represent the value read from F1 by the r.v. $Y_1$ and the value read from F2 by $Y_2$. Assume that the stored bit is represented by $X$, where $X$ is equally likely to be 0 or 1, barring any other information.

(a) Assume that $p_1 = 0.1$ and $p_2 = 0.2$, that is, the probability of flipping is low. Are $Y_1$ and $Y_2$ dependent? Explain using the definition of independence of random variables.
(b) Repeat when $p_1 = 0.5$ and $p_2 = 0.2$. Now are $Y_1$ and $Y_2$ dependent?
(c) Repeat when $p_1 = 0.7$ and $p_2 = 0.8$. Now are $Y_1$ and $Y_2$ dependent?
(d) For what values of $p_1$ and $p_2$ do you conjecture that $Y_1$ and $Y_2$ are dependent? Why do you think this is?

3.13 **Correlated basketball**
A basketball player attempts a shot and makes it. She attempts another shot and misses it. Her subsequent shots have success probability based on the proportion of her previous successful shots. What's the probability she makes 50 out of 100 shots? [Hint: Try looking for a pattern.]

3.14 **How to find a mate**
Imagine that there are $n$ people in the world. You want to find the best spouse. You date one person at a time. After dating a person, you must decide if you want to marry them. If you decide to marry, then you're done. If you decide not to marry, then that person will never again agree to marry you (they're on the "burn list"), and you move on to the next person.
Suppose that after dating a person you can accurately rank them in comparison with all the other people you've dated so far. You do not, however, know their rank relative to people you haven't dated. So, for example, you might early on date the person who is the best of the $n$, but you don't know that – you only know that this person is better than the people you've dated so far.
For the purpose of this problem, assume that each candidate has a unique score, uniformly distributed between 0 and 1. Your goal is to find the candidate with the highest score.

**Algorithm 3.9 (Marriage algorithm)**
1. *Date $r \ll n$ people. Rank those $r$ to determine the "best of $r$."*
2. *Now keep dating people until you find a person who is better than that "best of $r$" person.*
3. *As soon as you find such a person, marry them. If you never find such a person, you'll stay unwed.*

What value of $r$ maximizes $\mathbf{P}\{\text{end up marrying the best of } n\}$? When using that $r$, what is the probability that you end up marrying the best person? In your analysis, feel free to assume that $n$ is large and thus $H_n \approx \ln(n)$, by (1.17).